

Exploring the influencing factors of COVID-19 vaccination willingness: A mixed-method study in China

Yue Su^{1,2}, Sijia Li^{1,2}, Feng Huang^{1,2}, Jia Xue^{3,*}, and Tingshao Zhu^{1,2,*}

¹ CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; suy@psych.ac.cn (Y.S.); lisj@psych.ac.cn (S.L.); huangf@psych.ac.cn (F.H.); tszhu@psych.ac.cn (T.Z.)

² Dept. of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

³ Factor-Inwentash Faculty of Social Work & Faculty of Information, University of Toronto, Toronto M5S 1A1, Canada; jia.xue@utoronto.ca (J.X.)

* Correspondence: tszhu@psych.ac.cn (T.Z.), jia.xue@utoronto.ca (J.X.)

Abstract

The outbreak of COVID-19 has caused enormous damage to individuals' normal life and society's development. Recently, with the emergence of novel coronavirus variants, people around the world are still under the attack of COVID-19. And vaccine uptake is considered one of the most effective methods to defend against this epidemic. However, many people hold hesitant attitudes towards the COVID-19 vaccine. For the vaccination promotion work, it is important to find out what factors would influence the intention of vaccination. With the aim of practical use, this study identifies two main themes of this factor structure, that is, non-manipulable factors and manipulable factors. This study uses semi-structured interviews to explore the influencing factors related to vaccination willingness. For the interview results analysis, this study employs a mixed-method, which combines qualitative thematic analysis and quantitative topic modeling. The thematic analysis identifies nine key factors and topic modeling identifies seven key factors. After gathering and discussing these key factors, this study confirmed ten key factors as final factors. More specifically, non-manipulable factors contain Effectiveness and safety of vaccines and Vaccine feasibility, and manipulable factors include Convenience of vaccine, Responsibility, External reference, Necessity and urgency of vaccination uptake, Perceived threat, Perceived benefits, Scientific and objective expression, and Policy requirement. This study investigates key influencing factors that would affect public vaccination intentions in China and provides specific guidelines for vaccination promotion for authorities and public health workers.

Keywords: influencing factor, vaccination intention, vaccine uptake, vaccination willingness, thematic analysis, semi-structured interview, topic model

Introduction

Since the outbreak of the COVID-19 epidemic in December 2019, almost every country in the world has suffered a lot during a long period. Despite the general public, professionals, and authorities having made huge efforts to fight COVID-19 and made multiple measures to prevent the transmission of the coronavirus, it is hard to take total control of this epidemic transmission till now. And at present, vaccination uptake is still considered one of the most efficient methods to protect

people from virus infection and build group immunity to deter epidemic spread.

However, many people hold hesitant attitudes towards the COVID-19 vaccine. Vaccine hesitancy is considered an important problem when promoting vaccination uptake. SAGE (The Strategic Advisory Group of Experts on Immunization), which is a working group of WHO (World Health Organization) defines vaccine hesitancy as “delay in acceptance or refusal of vaccines despite availability of vaccine services. Vaccine hesitancy is complex and context-specific, varying across time, place, and vaccines” (The SAGE Vaccine Hesitancy Working Group, 2014).

Recently, many researchers have focused on COVID-19-related vaccine hesitancy and explored what factors could influence the intentions of the vaccine. Age, gender, and occupation are found to be associated with vaccine hesitancy (Wake, 2021). Besides, more in-depth research has been conducted. Wang and colleagues (2021) use an online questionnaire to investigate COVID-19 vaccination uptake intentions and influencing factors among overseas and domestic university students. They find that perceived benefits, perceived barriers, and cues to action were important influencing factors related to COVID-19 vaccination intention. In addition to a questionnaire, an interview is also a useful approach for a deep and specific investigation of how people think when talking about COVID-19 vaccination uptake. Mugur and colleagues (2021) conducted interviews with women leaving jails and find that for vaccine promotion, health education needs to be strengthened, and mistrust, misinformation, and conspiracy theories need to be decreased (Geana, Anderson, & Ramaswamy, 2021). And a qualitative study based on semi-structured interviews explores the perceptions of and hesitancy toward COVID-19 vaccination in older Chinese adults in Hong Kong and identifies barriers to vaccination and motivations for vaccination as two main themes (Siu, Cao, & Shum, 2022).

It is hard for researchers to reach an agreement on the influencing factors related to vaccination intention. This may be associated with the differences in participants, periods, and analysis methods. Focusing on the analysis methods, we intend to employ a mixed-method to make a better and deeper analysis of the interview results, which includes the traditional qualitative method as well as the quantitative machine learning method. The first analysis method is based on thematic analysis and manual coding. The other was with the application of topic modeling algorithms.

As we all know, the analysis of interview results mainly concentrates on manual coding. Researchers need to read and discuss transcripts carefully to make a consensus. However, new methods have emerged to analyze text contents in recent years, and topic modeling is a pretty popular approach in text analysis (Blei, Ng, & Jordan, 2003; Steyvers & Griffiths, 2006). Topic modeling is an unsupervised machine learning method and could provide multiple word clusters as the main topics for the understanding of the basic and underlying dimensions of linguistic data (Atkins et al., 2012). For example, Xue and colleagues use topic modeling to analyze tweets related to COVID-19 with the aim of public discourse investigation and identified 10 key themes of these tweets (Xue et al., 2020). Except for social media texts, interview transcripts are also appropriate materials for topic modeling, and we confirm the use of topic modeling in the couple-therapy trial as well (Atkins et al., 2012). However, as far as we know, topic modeling has not been used in the COVID-19 related interview study. In our study, we intend to use topic modeling as a supplementary and quantitative method for text data analysis. By the combination of manual coding and topic modeling, we could provide a novel and comprehensive perspective for the interview transcript analysis.

Our study aims to identify the influencing factors related to COVID-19 vaccine willingness in

China and provide references for vaccination promotion for authorities and public health workers. From a practical perspective, we want to focus on the factors that are of great practical value in vaccination promotion work. Thus, we have identified two main themes of the influencing factor structure, one is non-manipulable factors, and the other is manipulable factors. Non-manipulable factors refer to the factors that are directly related to the COVID-19 vaccine itself, such as experiment results and objective data. And thus, the core contents of these factors are hard to change and should keep the same no matter where or how these contents are published, such as vaccination rate. And manipulable factors refer to the factors that are indirectly related to the COVID-19 vaccine and commonly focus on outside circumstances rather than vaccine information per se. These factors are usually easy to manipulate in practice to influence the feelings of individuals.

Study 1

Method

We used semi-structured interviews to explore the influencing factors related to the COVID-19 vaccine. We aimed to investigate what people were concerned about the COVID-19 vaccine and what information would drive people to get vaccinated more actively. The interview guide was developed by four psychological experts (two teachers in the department of psychology and two students majoring in psychology) and they agreed on this interview guide after discussion and analysis. This interview guide contained two sections, one was about the concerns that people may have about a vaccine, and the other was about the persuasive information and influencing factors that people raised.

Participant recruitment

The participants were recruited by purposive and snowball sampling. By this sampling method, we aimed to recruit participants to maximize the diversity in age, gender, educational level, and geographical location. More specifically, the participant recruitment advertisements were posted on social media, and people who were interested in this interview would fill out the questionnaire contained in the advertisement and leave their phone number and available time. And we would select interested participants to interview to enrich the demographic information of our data sample. The participants were compensated with a cash reward as a thank for their time and contribution. The reward was paid based on the interview time and there was also a bonus according to the performance of participants, that is, 2 Yuan per minute and 0-30 Yuan bonus. Our study was approved in advance by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences (approved number: H15009).

Data collection

There were six trained students majoring in psychology involved in this experiment and interviews were carried out by one of these interviewers. All interviews were performed on Tencent Meeting, which was a remote meeting application providing easy-to-use and reliable cloud video and audio-conferencing services. These interviews lasted approximately 30-60 minutes. Informed consent was obtained by each participant. And interviews were audio-recorded for further analysis. The data collection ended and the sample size was confirmed when the data acquired from interviews reached a saturation point and there were no novel ideas arising (Marbán-Castro et al.,

2021).

Data analysis

The audio recordings were transcribed verbatim by the interviewers. Then, researchers performed thematic analysis to code interview data manually. In this research, we followed the thematic analysis steps proposed by Braun and Clarke (Bell et al., 2020) which included familiarizing ourselves with our data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report (Braun, & Clarke, 2006). To enhance the rigor of the analysis, the initial code development and the subsequent theme generation and definition were discussed between two researchers (the first author and the second author) and reached a final agreement (Bell et al., 2020). Thus, the codes and themes were confirmed. We also selected concrete examples for each theme.

Result

Demographic information

12 participants were included in our study. Table 1 showed the demographic information of our interview participants. The mean age of participants was 25.83, and the standard deviation was 2.98. To be more specific, the maximum age was 34, and the minimum was 23. Among these participants, 6 were male and 6 were female, which was a balanced gender ratio. As for education level, we found that one participant had a college diploma and eleven participants were bachelor's degree and above. The geological distribution of our participants was rather diverse. 5 participants were located in eastern China, 4 were situated in central China, and 3 were located in western China.

Table 1. Demographic information of interview participants

Characteristics	N=12
Age (Mean \pm SD)	25.83 \pm 2.98
Gender n (%)	
Male	6 (50%)
Female	6 (50%)
Education level n (%)	
High school (including technical secondary school) and below	0 (0%)
College (three-year or two-year college diploma)	1 (8.33%)
Bachelors and above	11 (91.67%)
Location	
Eastern China	5 (41.67%)
Central China	4 (33.33%)
Western China	3 (25.00%)

Interview result

After finishing all interviews, we analyzed interview transcripts. 31 initial codes were generated, please refer to more details in Appendix. Then, these 31 codes were grouped into 9 key factors, and these 9 factors were grouped into 2 main themes. The final interview results were shown in Table 2.

Theme 1: Non-manipulable factor

One key factor was divided into non-manipulable factors, that is, the effectiveness and safety of vaccines. This factor meant that the vaccine would not pose a threat to personal safety. More

specifically, this factor emphasized the detailed and clear, scientific, and rigorous explanations of vaccination uptake conditions, application groups of vaccination, vaccination restrictions, and possible adverse reactions. For example, participants indicated that they were pretty worried about the side effects after vaccination. And some participants wanted to know more about why some people were not suitable for vaccination.

Theme 2: Manipulable factors

Manipulable factors contained 8 key factors, including the convenience of vaccine, responsibility, external reference, necessity and urgency of vaccination uptake, perceived threat, perceived benefits, scientific and objective expression, and policy requirement.

The convenience of vaccines referred to the costs that individuals need to consider and pay in the process of vaccination, such as distance, and time. Participants mentioned that if there were lots of vaccination sites and it was convenient to get vaccinated, they would be more willing to get the COVID-19 vaccine.

Responsibility meant that people hold the awareness, emotion, and belief that an individual should be responsible for himself and others, family and group, country and society. Moreover, they also have the corresponding conscious attitude that an individual should abide by norms, accept responsibility, and fulfill obligations. Some participants indicated that vaccination uptake is a sign of being responsible to others and society.

External reference meant that when an individual intended to take a certain behavior or form a certain attitude, he would use a specific individual or group as the reference and comparison object. These reference objects included the specific social group, the majority of people in the society, people around him or her, and people with different political and cultural backgrounds. Some participants mentioned that how people around them see vaccines were important for them. And if near friends and family chose to get vaccinated, these participants would also approve of the vaccine.

The necessity and urgency of vaccination uptake emphasized the intensity and urgency degree of the need for vaccination that individuals felt under the current social and epidemic situation. Several participants mentioned that news reports that the domestic confirmed cases began to grow would increase their intention of vaccination.

Perceived threat referred to people feeling some threats to freedom when some acts with the aim of persuasion interfere with individual free choice. Some participants indicated that lots of reports said people must get a vaccination instead of should get a vaccination, that kind of expression made them hard to accept the COVID-19 vaccine.

Perceived benefits meant benefits perceived by individuals after vaccination. For example, participants mentioned that they would get vaccinated actively if it was easier to go in and out of campus.

Scientific and objective expression emphasized that the vaccine-related information should avoid containing arbitrary or ambiguous expressions, provide pros and cons of vaccination from an objective perspective, and use accurate and easy-to-read language. Some participants mentioned that they heard the effectiveness rate of COVID-19 vaccines in Brazil was not high. And they thought it was pretty honest to publish such information instead of mere praise of vaccine.

Policy requirements meant that an individual would abandon his or her own opinion and behave by policy demands, and group regulations. Some participants indicated that if the authority authorized mandatory vaccination against COVID-19, they would obey this requirement and get vaccinated.

Table 2. Key factors and codes

The main theme, the key factor, and codes
Non-manipulable factor
<i>Effectiveness and safety of vaccines</i>
Vaccine safety, The restriction of vaccination uptake, Introduction of vaccine general information, Specific explanation of vaccination requirement, Data exhibition, Vaccine effectiveness, Providing a comparison
Manipulable factors
<i>Convenience of vaccine</i>
Convenience of vaccine
<i>Responsibility</i>
Responsibility
<i>External reference</i>
Group identification, Conformity, Reference group, In-group pressure, Psychological distance, Vicarious experience, Specialization, Credibility. Attraction, Official position, Unofficial position
<i>Necessity and urgency of vaccination uptake</i>
Risk perception, Scarcity, Deadline strategy
<i>Perceived threat</i>
Perceived threat
<i>Perceived benefits</i>
Vaccination certification, Personal involvement degree, Extra benefits
<i>Scientific and objective expression</i>
Dual role persuasion, Certainty of vaccination information, Popular science
<i>Policy requirement</i>
Policy requirement

Study 2

Method

Based on the transcripts obtained in Study 1, we collected all sets of words spoken by participants. More specifically, we defined a set of words spoken by a certain participant as one document and we got 12 documents in total.

Before conducting a topic model, we used Python to preprocess all documents to ensure the quality and interpretability of analysis results. The preprocessing steps include the following three steps:

1. We removed all special symbols except for English characters, Chinese characters, and numbers. It should be noted that all interviews were conducted in Chinese and all documents were written in Chinese. But considering some English-related expressions, such as mRNA, also appeared in interviews, we kept English characters as well.
2. We used Jieba to conduct word segmentation. Jieba was a Python package and was commonly

used in Chinese segmentation. In our analysis, Jieba was called in Python to separate sentences into words and phrases. Considering new terminologies and expressions are being developed constantly as the world changes, we added some specific terms into the built-in user dictionary in Jieba, such as Bilibili, Health Code, and Li Jiaqi.

3. We removed stop words and words whose length was less than 2 characters in the segmentation results of documents. This word elimination could reduce the appearance of irrelevant terms and noises and enhance the performance of the topic model algorithm (Calderón, de la Vega, & Herrero, 2020).

We got the final dataset which was ready for topic modeling. Then, we used Latent Dirichlet Allocation (LDA) algorithm to explore the main topics and structures of the dataset. LDA was a common topic modeling method and was frequently used in the investigation of multiple topics in a set of documents (Ramage et al., 2009; Xue et al., 2020). LDA was a probabilistic model and could infer latent thematic structures from documents without prior knowledge or manual labels (Calderón, de la Vega, & Herrero, 2020). The Mallet version of LDA was implemented in the Python package Gensim and was considered to perform better than LDA (Wang & Li, 2020). Therefore, we used the Mallet version of LDA to build the topic model and investigate the thematic patterns of documents.

The number of topics was an important parameter for the model performance. To determine the appropriate number of topics, we tried different numbers of topics that varied from 2 to 30 and calculated the coherence score for each model. A coherence score was advisable to be used in the measurement of model performance. The higher coherence score indicated a more precise model (Calderón, de la Vega, & Herrero, 2020). Based on the coherence score, the most appropriate number of topics was chosen and the final model of topics was confirmed.

In the next step, we employed pyLDAvis to visualize the topic model result. pyLDAvis was a Python package providing the visual interface for inter-topic distance (Chuang et al., 2012). Each bubble in the inter-topic distance map represented a topic and meanwhile, the area of the bubble meant the prevalence of the topic (Hariman, Vries, & Smeets, 2018). And the distances of topics were multidimensionally scaled onto two axes (Sievert & Shirley, 2014). The relatively big and non-overlapping bubbles in the map would suggest a proper topic model result (Hariman, Vries, & Smeets, 2018).

The last step of topic modeling was to analyze, identify and describe themes according to the results of the LDA algorithm. After identifying the number of topics in LDA, two authors (Y.S and S.L) discussed the top 20 keywords of each topic and reviewed the corresponding expressions in interview transcripts. Then, they reached an agreement on the label and description of each theme.

Result

To determine the appropriate number of topics for topic modeling, we calculated the coherence score for each topic model. Figure 1 presented the coherence score result for each topic model, which was built on a different number of topics. We finally chose 7 as the most appropriate number of topics for topic modeling based on two considerations. First, when the number of the topics was as 7, the coherence scores were at a high level indicating the good performance of the topic model. Second, our corpus was not very large, and too much number of topics would dilute the main focus of our text materials. Thus, in comparison with other numbers of topics with rather high coherence scores, such as 16 and 18, number 7 contained the important topics as much as possible, on the one

hand, and avoided the risk of confusing the main focus, on the other (Lyu, Le Han, & Luli, 2021).

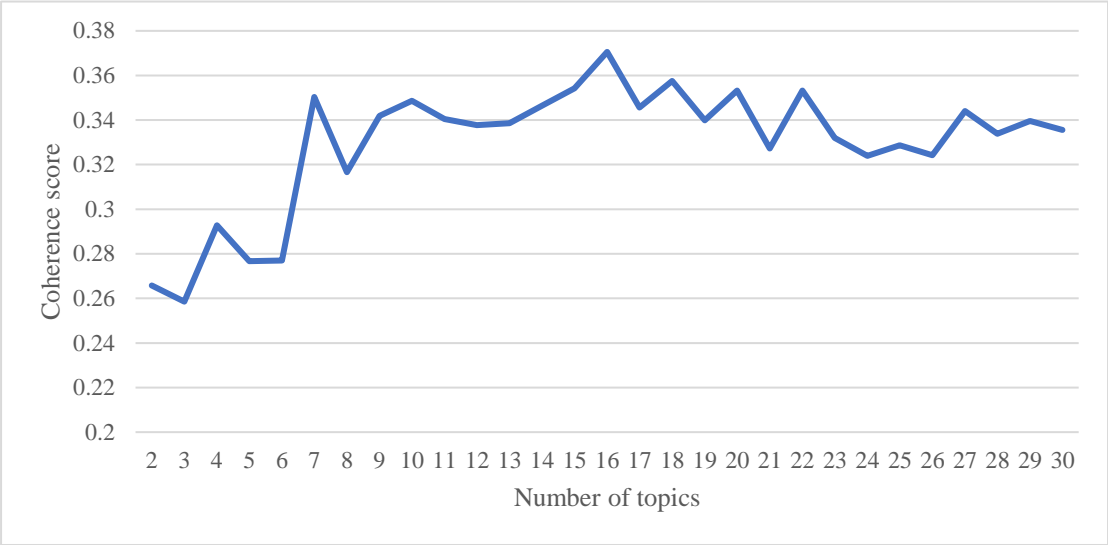


Figure 1. Coherence score for each topic model built on a different number of topics

The inter-topic distance map was shown in Figure 2. Each bubble represented a topic from Topic 1 to Topic 7 in our study. And the bubbles in Figure 2 were relatively big and not overlapped, which suggested a good topic model and confirmed the reasonability of topic number 7. The topics are ordered in descending order of prevalence in the corpus (Hariman, Vries, & Smeets, 2018).

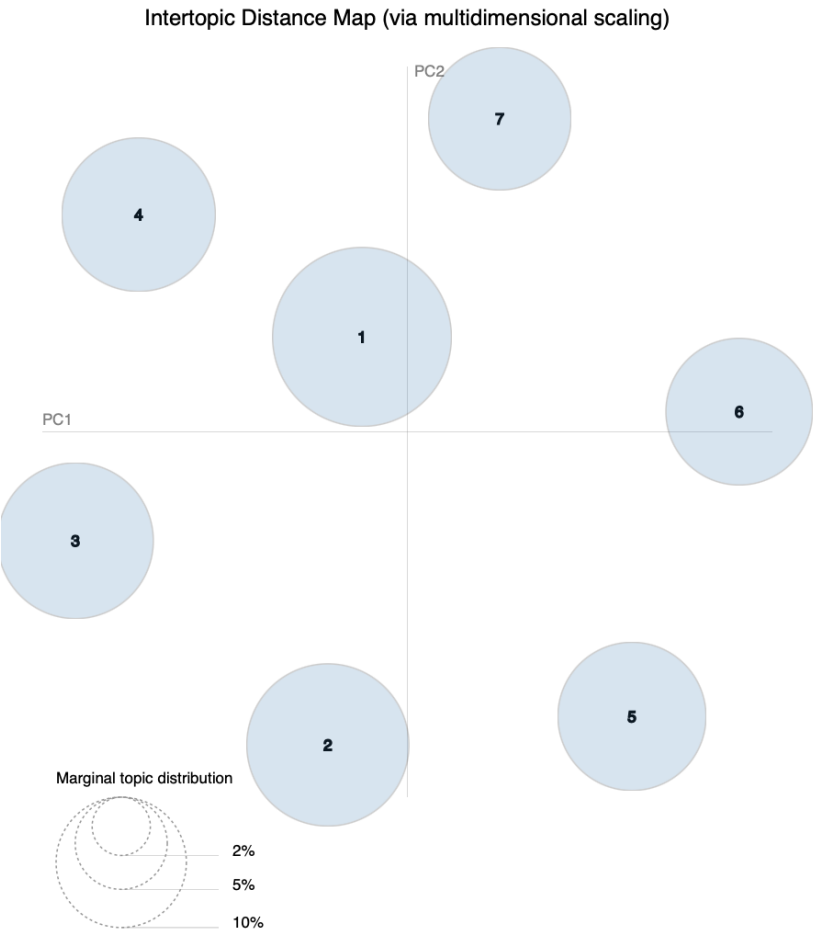


Figure 2. Inter-topic distance map

Table 1 showed the result of the topic models with 7 topics and the top 20 relevant words of each topic in the corpus. The numbers of topics were by the numbers shown in the inter-topic distance map.

Table 1. the topic model result with 7 topics and the relevant words

Topic	The top 20 relevant word	% of tokens
1	Vaccine, get vaccinated, nation, attitude, around, information, promote, COVID-19, lots of, friend, matter, others, basically, really, job, official, after all, China, wait-and-see, several	18.8
2	Vaccine, that sort of, school, a little, popular science, community, two-shot, WeChat Moments, do not get vaccinated, is there any, influence, reason, time, first shot, second shot, three-shot, inconvenient, social, one-shot, life	15.5
3	Get vaccinated, side effect, school, thing, classmate, do not want to, matter, do not get vaccinated, body, suggestion, doctor, video, mentality, definitely, in this regard, six months, a few days, domestic, willingness, fear	14.3
4	Get vaccinated, definitely, influence, COVID-19 vaccine, adverse reactions, nation, concerns, a crowd of people, if, worry, epidemic, place, nucleic acid, a lot of, infect, allergy, child, impression, on earth, on the Internet	13.9
5	Information, influence, focus on, policy, particularly, Weibo, I will, said to be, publicity, research and development, thing, precautions, like this, sometimes, effect, increase, remember, because of, observe, video	12.9
6	feel like, a little, news, kind of, epidemic, for, concerns, popular, situation, accept, particularly, time, infect, not quite, antibody, free, national people, less than, TikTok, effect	12.7
7	Get vaccinated, vaccine, media, America, lots of, safety, probably, reports, temporary, be from, factors, data, scientific, appointment, tell, authority, indeed, this piece of, risk, the vaccination rate	12

After analysis and discussion of the seven topics grouped by topic modeling, we identified these 7 topics as our final topic categories, labeled each topic, and grouped these seven topics into two themes, which was by the theme structure that we suggested before. This result was shown in Table 2.

Table 2. Seven topics produced by LDA with labeled names and examples

Theme and topics	Examples
Theme 1: Non-manipulable factors	
Topic #3 <i>Side effects of the COVID-19 vaccine</i>	Some people, I think, maybe a little scarred psychologically. They are uncertain about this vaccine. Because when we published the information about this vaccine, we mentioned that this vaccine will cause some unpleasant symptoms, such as skin redness, fever, fatigue, nausea, and headaches, and then they might feel a little scared. (Interview #10)

		<p>Another point may be related to doctors, such as the video channels or TikTok accounts of doctors. If doctors post videos to further explain the effects of the COVID-19 vaccine, as well as side effects and adverse reactions, and if they demonstrate such information very clearly, I think I will definitely be willing to get vaccinated. (Interview #5)</p> <p>There are some individuals with uncommon constitutions, such as suffering from allergies and purpura before. Can this group, on earth, get this vaccine? Some people asked me, "My child has had purpura before, can he get vaccinated? His school asks him to get vaccinated, can he?" Then I'm not very clear about this. (Interview #8)</p> <p>In fact, I still don't understand why we need to do a nucleic acid test once a week even after vaccine uptake? All our hospitalized patients need nucleic acid tests. But some family members did not cooperate with us. They claimed that they have been vaccinated, so they didn't need a nucleic acid test. (Interview #8)</p>
Topic #4 <i>Vaccine feasibility</i>		
Theme 2: Manipulable factors		
Topic #1 <i>The influences from others' experiences</i>	#2	<p>And the feelings of the people around me about being vaccinated will be more convincing for me. (Interview #2)</p> <p>(I) want to wait first and see the reactions of whom have already gotten vaccinated. (Interview #6)</p> <p>Even if he goes to the vaccination site for vaccination, it is still possible to fail to get vaccinated. And he has to wait for the second shot after getting the first shot, and sometimes the third shot is also required. He may feel it is not worthy to cost such time. (Interview #4)</p> <p>There is a piece of news, for example, that vaccines will be delivered to your home or your community. And time to get vaccinated is within a reasonable period, such as just going downstairs or walking for a few minutes. The vaccination process is very fast and convenient. Such a situation can also make me more willing (to get vaccinated). (Interview #11)</p> <p>If I can see a very authoritative popular science document that can make people like me who are not willing to think about things easy to understand, then I would feel I understand it and consider getting vaccinated. Especially for the distinctions among different types of vaccines, the period that vaccines are effective, and the precautions of vaccination, the popular science document should explain these aspects clearly and concisely. (Interview #4)</p> <p>Take the lecture that I just mentioned as an example the content of the lecture is to popularize why this vaccine can be developed in such a short period and it runs ahead many steps compared to other countries. If these questions could be popularized in a layman-friendly way, I would be more willing to get vaccinated. Because now everyone only knows that the vaccine has been developed and is available for the public, but they do not know how it was developed in fact. (Interview #7)</p>
Topic #5 <i>Direct and transparent information</i>		
Topic #6 <i>Positive perception of</i>		<p>In this area where I live, the epidemic is not very serious, and the protection work is very good. When the epidemic was the worst, people in our area were not infected. After the vaccine is available, it does not matter for many people to get</p>

<i>epidemic situation</i>	vaccinated. As long as we don't go out, we will be fine. Many people think so. After all, people hold the opinion that we were all safe in this area when the epidemic was so severe, and now the epidemic has passed for so long, (Interview #3) In the case that everyone is vaccinated, he can also be considered safe. That is, if everyone could block the chain of transmission, then he doesn't need to get vaccinated. (Interview #9)
Topic #7 <i>Safe and reliable information related to the COVID-19 vaccine</i>	I think the official media, academicians, such as Nanshan Zhong, and the remarks or appeals made by these pretty authoritative people would have a very positive impression on me. (Interview #1) I think it is more customary for me to read detailed parameters and factual reports. I don't believe it if you simply tell me how safe this thing is. Do not say conclusions, give me data. (Interview #12)

Theme 1: Non-manipulable factors

This theme contained two topics, and these two topics both focused on the information which was directly related to the COVID-19 vaccine itself. The first topic (Topic #3) was named “Side effects of COVID-19 vaccine”. This topic concentrated on the concerns about the side effects of the vaccines. More specifically, how their body would feel when people get vaccinated makes them worried. The second topic (Topic #4) was labeled “Vaccine feasibility”. This topic referred to the explanation of the feasibility range of vaccines, that is, what kind of people could get vaccinated, what kind of people should wait, what kind of people should deny vaccination, and in what situation vaccination would be useful and on what situation full vaccination would not be enough, and other epidemic protection approaches are also required (e. g. nucleic acid test).

Theme 2: Manipulable factors

This theme contained five topics, and these topics mainly focused on the influences of external factors instead of vaccine information. The first topic (Topic #1) was named “The influences from others’ experiences”. This topic meant that people would refer to the experiences and behaviors of other people when deciding on vaccination uptakes, such as friends and colleagues. The feedback of others was of great value to them. The second topic (Topic #2) was labeled “The potential costs during the vaccination process”. This topic included two aspects. One was the social pressure that people might occur if they would not get vaccinated. The other was the costs of time and energy that people should pay if they would consider getting vaccinated. In a word, they were all the costs that people should consider when they made a plan about vaccination. The name of the third topic (Topic #5) was “Direct and transparent information”. This topic emphasized that the language and expression related to the vaccine should be unambiguous and direct, sincere and transparent. The fourth topic (Topic #6) was named “Positive perception of the epidemic situation”. This topic referred to the necessity of vaccination uptake, and when people felt positive about living circumstances and thought the infection risk was low, people might be more unwilling to get vaccinated. The fifth topic (Topic #7) was labeled “Safe and reliable information related to COVID-19 vaccine”. This topic showed people’s attention to trustworthy information related to vaccine safety. More specifically, official media, medical professionals, and scientific reports that support and verify vaccine safety would have some impact on some people.

Discussion

This study explored influencing factors related to public vaccination willingness based on semi-structured interviews. We recruited 12 participants in our experiment and obtained verbatim transcripts of each interview. The first study used thematic analysis to code transcripts manually. The thematic analysis identified one key factor in the theme “non-manipulable factors”, and eight key factors in the theme “manipulable factors”. The second study employed the topic modeling method LDA to examine the semantic pattern of transcripts. The topic modeling confirmed two key factors in the theme “non-manipulable factors”, and five key factors in the theme “manipulable factors”.

The mixed analysis methods provided two similar but also subtly different results. For the theme “non-manipulable factors”, the thematic analysis identified one key factor, the Effectiveness, and safety of vaccines. The topic model identified two key factors, Side effects of the COVID-19 vaccine and Vaccine feasibility. We found that topic modeling confirmed more factors than thematic analysis. And it was obvious that the safety contained in the effectiveness and safety of vaccines focused on the side effects that vaccines may cause, which was similar to the Side effects of the COVID-19 vaccine. However, Vaccine feasibility emphasized the situation and group that vaccines should and should not apply. Such factor found the semantic connection between the targeted group and useful situation, which was different from the concentration on effectiveness and safety of vaccines and indeed supplemented a new perspective of non-manipulable factors.

For the theme “Manipulable factors”, the thematic analysis identified eight key factors, that is, Convenience of vaccine, Responsibility, External reference, Necessity and urgency of vaccination uptake, Perceived threat, Perceived benefits, Scientific and objective expression, Policy requirements. Topic model identified five key factors, which were The influences from others’ experiences, The potential costs during the vaccination process, Direct and transparent information, Positive perception of epidemic situation, Safe and reliable information related to the COVID-19 vaccine.

According to the meanings of these factors discussed above, we found that External reference was pretty similar to The influences from others’ experiences, and Scientific and objective expression was similar to Direct and transparent information. Besides, the Convenience of the vaccine was partly similar to The potential costs during the vaccination process, because the latter also mentioned the social pressure related to vaccination, which was included in the External reference in the thematic analysis. Moreover, Safe and reliable information related to the COVID-19 vaccine in topic modeling was also included in external reference in thematic analysis. In addition, one interesting finding was one factor that both included in two methods was presented in different directions. The necessity and urgency of vaccination uptake focused on the pressure and risk from situations make people feel urgent and be more active to get the vaccination. Meanwhile, Positive perception of an epidemic situation concentrated on the positive feeling of the epidemic would decrease the public’s willingness to vaccination. Nevertheless, these two factors both emphasize the necessity of vaccine uptake. Besides, the thematic analysis also found Responsibility, Perceived threat, Perceived benefits, and Policy requirements, which were not included in topic modeling. Therefore, these findings suggested that for manipulable factors, the thematic analysis provided more detailed and distinct results than topic modeling.

Above all, after taking both thematic analysis and topic modeling into consideration, we added Vaccine feasibility to non-manipulable factors based on our thematic analysis results, and finally identified ten key factors related to COVID-19 vaccination intention.

Some factors in this study were also identified in the previous study. In a systematic review written by Wake (2021), perceived risk of COVID-19, norms, perceived benefit of vaccine, perceived vaccine barriers, perceived efficacy of the COVID-19 vaccination, COVID-19 vaccine safety concerns were considered to be associated with the willingness of COVID-19 vaccine uptake (Wake, 2021). However, as far as we know, this is the first study to take Vaccine feasibility, Responsibility, and Policy requirement as distinct key factors and incorporate them into a complete factor structure, which could provide diverse perspectives to discuss how to increase public vaccination intention.

This study used a mixed-method to analyze interview results. Our findings suggested that it is feasible to use thematic analysis with the supplementation of topic modeling. In most cases, the thematic analysis could provide detailed insights. In contrast, the topics obtained by the topic model were relatively rough. However, topic modeling could find some novel relationships ignored by people sometimes.

There were some developed models to explain influencing factors related to vaccination willingness. SAGE working group proposed the 3C model including confidence, convenience, and complacency (MacDonald, 2015). Thomson et al. (2016) suggested the 5A model including acceptance, access, awareness, affordability, and activation (Thomson, Robinson, & Vallée-Tourangeau, 2016). These models all summarized influencing factors into a mature system and formed a good theory. Different from these models, this study focused on the practical value and framed our factor structure into two themes, that is, non-manipulable factors and manipulable factors. As far as we know, this was the first study to propose one influencing factor structure related to the COVID-19 vaccine based on practical application. This structure could provide a clear view of how these factors should be used in the real world.

Our findings deepened the understanding of related factors that would influence the public COVID-19 vaccine uptake intention in China and contributed to the guideline of targeted vaccination promotion programs. Based on key factors proposed in this study, authorities and public health workers could focus on one or several factors according to the specific social situation and epidemic condition. For example, if the COVID-19 outbreak started in a country near our country, authorities could emphasize the severity and risk this country faced. Then, people may feel they were still not far from COVID-19 and their awareness of vaccination would be raised.

Conclusion

This study aimed to explore the influencing factors related to COVID-19 vaccination intention in China. And we used semi-structured interviews and employed mixed methods to analyze interview results, including thematic analysis and topic modeling. Results suggested thematic analysis could produce a more detailed analysis than topic modeling in general. However, topic modeling could examine novel relationships among expressions of participants and offer a new perspective. In conclusion, we identified ten key factors in our study. More specifically, the Effectiveness and safety of vaccines and Vaccine feasibility were divided into non-manipulable factors, Convenience of vaccine, Responsibility, External reference, Necessity and urgency of vaccination uptake, Perceived

threat, Perceived benefits, Scientific and objective expression, and Policy requirements were divided into manipulable factors. We suggested authorities and public health workers could focus on key factors proposed in this study and carry out a targeted vaccination promotion program.

Reference

Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R., & Christensen, A. (2012). Topic models: a novel method for modeling couple and family text data. *Journal of family psychology*, 26(5), 816.

Bell, S., Clarke, R., Mounier-Jack, S., Walker, J. L., & Paterson, P. (2020). Parents' and guardians' views on the acceptability of a future COVID-19 vaccine: A multi-methods study in England. *Vaccine*, 38(49), 7789-7798.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.

Calderón, C. A., de la Vega, G., & Herrero, D. B. (2020). Topic modeling and characterization of hate speech against immigrants on Twitter around the emergence of a far-right party in Spain. *Social Sciences*, 9(11), 188.

Chuang, J., Ramage, D., Manning, C., & Heer, J. (2012, May). Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 443-452).

Geana, M. V., Anderson, S., & Ramaswamy, M. (2021). COVID-19 vaccine hesitancy among women leaving jails: A qualitative study. *Public Health Nursing*, 38(5), 892-896.

Hariman, N., Vries, M. D., & Smeets, I. (2018, November). Topic modeling for exploring cancer-related coverage in journalistic texts. In *Benelux conference on artificial intelligence* (pp. 43-51). Springer, Cham.

Lyu, J. C., Le Han, E., & Luli, G. K. (2021). COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. *Journal of medical Internet research*, 23(6), e24435.

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161-4164.

Marbán-Castro, E., Pons-Duran, C., García-Otero, L., Chen, H., Herrera, L. B., Gil, M. D. M., ... & Enguita-Fernández, C. (2021). Acceptability of clinical trials on COVID-19 during pregnancy among pregnant women and healthcare providers: a qualitative study. *International journal of environmental research and public health*, 18(20), 10717.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009, August). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 248-256).

Sievert, C., & Shirley, K. (2014, June). LDavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70).

Siu, J. Y. M., Cao, Y., & Shum, D. H. (2022). Perceptions of and hesitancy toward COVID-19 vaccination in older Chinese adults in Hong Kong: a qualitative study. *BMC geriatrics*, 22(1), 1-16.

Steyvers, M., & Griffiths, T. (2006). Probabilistic topic models. In T. Landauer, D. McNamara,

S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning*, (pp. 427-448). Mahwah, NJ: Erlbaum.

The SAGE Vaccine Hesitancy Working Group. Report of the SAGE working group on vaccine hesitancy. 2014;

http://www.who.int/immunization/sage/sage_wg_vaccine_hesitancy_apr12/en/index.html

Thomson, A., Robinson, K., & Vallée-Tourangeau, G. (2016). The 5As: A practical taxonomy for the determinants of vaccine uptake. *Vaccine*, 34(8), 1018-1024.

Wake, A. D. (2021). The willingness to receive COVID-19 vaccine and its associated factors: "vaccination refusal could prolong the war of this pandemic"—a systematic review. *Risk management and healthcare policy*, 14, 2609.

Wang, J., & Li, C. (2020, July). Research on User Satisfaction of Video Education Application Based on Reviews. In *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)* (pp. 340-343). IEEE.

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PloS one*, 15(9), e0239441.

Appendix. 31 Codes generated from the interview transcripts

No.	Code name	Operational definition
1	Vaccine safety	The vaccine does not pose a threat to personal safety, such as a low incidence rate of serious adverse reactions, and the good health condition of population after vaccination
2	Vaccine effectiveness	The effective result and effective time period. That is, vaccine could protect individuals from the infection of corona virus with a high probability, and vaccine is effective in a specific period of time.
3	Convenience of vaccine	Costs that need to be considered in the process of vaccination, such as distance, time.
4	The restriction of vaccination uptake	The specific groups that are not suitable for vaccination
5	Introduction of vaccine general information	Basic information about vaccine, such as the components of vaccine.
6	Specific explanation of vaccination requirement	Explanation of specific behaviors and requirements in the process of vaccine promotion, such as the specific reasons for people who are not suitable for vaccination
7	Unofficial position	The publisher of the information is not the official media, but some medias with a neutral political position and no political purpose.
8	Official position	The confidence in public power.
9	Scarcity	The scarcer the product, the more people want it. This means that people are more likely to desire or demand products if they are told they are hard to get.
10	Data exhibition	Provide specific numeric values in the expression.
11	Popular science	Provide vaccine-related scientific information for public in

		accurate, rigorous, and easy-to-read language.
12	Responsibility	People hold the awareness, emotion and belief that an individual should be responsible for himself and others, family and group, country and society. And they also have the corresponding conscious attitude that an individual should abide by norms, accept responsibility, and fulfill obligations.
13	Group identification	The individual realizes that he belongs to a particular social group, and at the same time recognizes the emotion and value that group brings to him. This group identification will form a kind of spiritual cohesion, which to some extent regulates the values and moral norms of group members.
14	Risk perception	Individuals' subjective judgments about how safe their current environment is, how serious the epidemic is, and how much their infection risk of is.
15	Conformity	An individual tends to follow group norms to form beliefs, and attitudes and take behaviors.
16	Policy requirement	An individual abandons his or her own opinion and behaves in accordance with policy demands, group regulations.
17	Specializations	Specialization means the characteristics of information disseminator, including professional ability, well-trained knowledge and rich experience.
18	Credibility	Trustworthiness means information receivers believe that the communicator is able and willing to provide objective, fair, true and effective information. Communicator should be sincere, honest and objective, and have no specific communication motives and intentions.
19	Attraction	Attraction means we are more likely to follow a request from someone we admire or like.
20	Dual role persuasion	The information of pros and cons should be both provided.
21	Perceived threat	Acts with the aim of persuasion. Such acts can be regarded as a threat to freedom if they interfere with individual free choice.
22	Deadline strategy	Increase compliance by indicating to individuals that they have limited time to obtain certain benefits.
23	Vaccination certification	People who take the COVID-19 vaccine will get certification in a certain form, such as the unique health code.
24	Personal involvement degree	Individuals' subjective experience of the relationship between the personal life and vaccination uptake.
25	Extra benefits	Additional benefits except for vaccine benefits per se.
26	Providing comparison	Provide different quality and different effects of goods for users to compare and choose.
27	Reference group	An individual or group who work as a reference when a person forms an attitude
28	Vicarious experience	Individuals can gain knowledge about vaccination by observing how others behave.

29	In-group pressure	An influence that a group has on its members. When a group member's thoughts or behaviors conflict with group norms, members feel a mental pressure that they need to abide by group norms in order to maintain their relationship with the group.
30	Certainty of vaccination information	Does not contain arbitrary or ambiguous expressions.
31	Psychological distance	The emotional or psychological distance that an individual maintains with another object.
